

MORE: Measurement and Correlation Based Variational Quantum Circuit for Multi-classification

Jindi Wu

Department of Computer Science
William & Mary
Williamsburg, VA, USA
jwu21@wm.edu

Tianjie Hu

Department of Computer Science
William & Mary
Williamsburg, VA, USA
thu04@wm.edu

Qun Li

Department of Computer Science
William & Mary
Williamsburg, VA, USA
liquan@cs.wm.edu

Abstract—Quantum computing has shown considerable promise for compute-intensive tasks in recent years. For instance, classification tasks based on quantum neural networks (QNN) have garnered significant interest from researchers and have been evaluated in various scenarios. However, the majority of quantum classifiers are currently limited to binary classification tasks due to either constrained quantum computing resources or the need for intensive classical post-processing. In this paper, we propose an efficient quantum multi-classifier called MORE, which stands for **m**easurement and **c**orrelation based variational quantum multi-classifier. MORE adopts the *same* variational ansatz as binary classifiers while performing multi-classification by fully utilizing the quantum information of a *single* readout qubit. To extract the complete information from the readout qubit, we select three observables that form the basis of a two-dimensional Hilbert space. We then use the quantum state tomography technique to reconstruct the readout state from the measurement results. Afterward, we explore the correlation between classes to determine the quantum labels for classes using the variational quantum clustering approach. Next, quantum label-based supervised learning is performed to identify the mapping between the input data and their corresponding quantum labels. Finally, the predicted label is determined by its closest quantum label when using the classifier. We implement this approach using the Qiskit Python library and evaluate it through extensive experiments on both noise-free and noisy quantum systems. Our evaluation results demonstrate that MORE, despite using a simple ansatz and limited quantum resources, achieves advanced performance.

Index Terms—quantum machine learning, quantum multi-classifier, variational quantum algorithm, hybrid quantum-classical method, observable-based method

I. INTRODUCTION

Quantum computing derived from quantum mechanics can exponentially accelerate the solution of specific problems compared to classical computing, because of the unique properties of superposition and entanglement [1]. Nevertheless, we are now in the Noisy Intermediate-Scale Quantum (NISQ) era, in which quantum machines contain just 50-100 qubits and are error-prone. Several months ago, IBM unveiled Osprey, the world’s largest quantum computer with 433 qubits [2]. This effort will provide the groundwork for the next era of quantum-centric supercomputing. However, its computing procedure is still plagued with qubit decoherence errors, gate errors, and measurement errors [3]. Quantum Machine learning (QML) is one of the most attractive applications of quantum computing

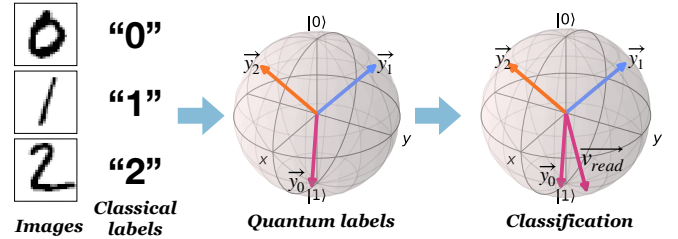


Fig. 1. Two-step MORE. First, the classical labels are converted into quantum labels in a two-dimensional Hilbert space based on the interclass correlation. Next, quantum label-based supervised learning is employed to minimize the difference between the readout state and its correct quantum label.

because of its demand for computing power and its resilience to noise [4]–[6]. In recent years, many efforts have been made to develop QML techniques as the counterparts of classical ones, such as quantum k-nearest neighbor [6], [7], quantum support vector machines [8]–[10], Quantum clustering [11], [12], and quantum neural networks (QNNs) [13]–[15]. Since classical neural networks have achieved remarkable success in various fields [16]–[20], QNNs have also been attempted for data processing in many domains, including finance [21], [22], chemistry [23], [24] and healthcare [25]–[29].

Classification is a fundamental data processing technique in many domains, making QNN-based classifiers an area of significant interest. In 2018, *Farhi and Neven* proposed the first QNN classifiers for near-term processors [13]. Following this, in 2019, *Cong et al.* proposed quantum convolutional neural networks (QCNNs) for image processing, which adopt the architecture design of its classical analog [30]. Based on these two quantum classifier designs, numerous variants are shown [31]–[34]. Moreover, some other QNN designs are suggested in [34]–[39].

Prior works on QNN-based classification have predominantly focused on binary classification tasks, with relatively few addressing multi-classification problems. This is partly due to the fact that QNNs typically produce binary measurement results, meaning that a single readout qubit can only represent two classes. The popular solutions for QNN-based multi-classification are: (1) use extra ancillary qubits to represent labels for multiple classes. However, qubits are a limited

resource on NISQ machines, and using more ancillary qubits reduces the number of qubits available for data processing, thus limiting the capacity of the QNN. (2) Select a subset of qubits at the end of the QNN circuit to serve as the readout for multi-class labels. However, if the one-hot encoding is used to represent labels, the number of classes a QNN-based classifier can handle is limited by the number of qubits available in the circuit. On the other hand, if binary encoding is used to represent labels, there may be cases where measurement results are not interpretable, which is a limitation that also applies to the previously mentioned solution. Specifically, the binary property of qubit measurement results enables an n -qubit state to include 2^n possible labels. Thus, given a p -classification task, where $p \in (2^{n-1}, 2^n)$, some output states will not map to any valid class. Especially when $p = 2^{n-1} + 1$, almost half of the label space is wasted. (3) Attaching classical fully-connected layers after QNNs is a common approach to mapping the QNN outcomes to multiple classes. However, this may incur high computational costs because the fully-connected layers contain a large number of parameters to train. Moreover, adding classical layers after QNNs may offset the potential quantum advantage of using QNNs for classification. And (4) decomposing the multi-classification task into several binary classification ones. The downsides include a complicated training process and lengthy training and inference times due to the need to train multiple binary classifiers. Therefore, there is a need for a novel QNN-based multi-classifier that features a concise variational quantum circuit, a streamlined classical post-processing procedure, and a straightforward training process.

In this article, we propose *MORE* approach, a novel QNN multi-classifier that utilizes measurements and interclass correlations. *MORE* is designed to be resource-efficient, featuring a simple ansatz similar to that used in binary classification, with only one readout qubit at the end of the circuit. *MORE* is a two-step approach, as shown in fig. 1. First, it converts the classical labels of training data into quantum labels, which are quantum states reconstructed from the measurement results of the readout qubit. The quantum label of each class is determined using the variational quantum clustering method that considers the correlation between classes. Then, quantum label-based supervised learning is conducted to converge the QNN and achieve the desired performance quickly. Compared to prior QNN-based multi-classifiers, *MORE* requires fewer qubits and quantum gates, resulting in fewer gate and decoherence errors during the quantum state evolution. Moreover, *MORE* is computationally efficient, with a fixed amount of quantum computation and linearly increasing classical computation, allowing scalability to more complex tasks.

The contribution of this work is fourfold:

- We present the attempt to design efficient quantum multi-classifiers by leveraging the full quantum information contained in a single qubit.
- We offer a technique for converting the classical labels of training data into quantum states by investigating interclass correlations.

- We introduce a quantum label-based quantum supervised learning approach that includes a loss adjuster to improve the quality of the QNN model.
- We implement the proposed *MORE* approach and evaluate it from many perspectives using comprehensive experiments. And our experiment results demonstrate the advantages of our proposed approach.

In the remaining content of this article, we provide some basic knowledge needed for understanding this article in section II. We then introduce our motivation for the proposed approach by discussing quantum state tomography, which is a method for reconstructing quantum states, in Section III. The proposed two-step approach *MORE* is detailed in section IV. The first step, variational quantum clustering based on interclass correlation, is described in subsection IV-B, then the second step, quantum label-based supervised learning, is discussed in subsection IV-C. Then, we present the evaluation setup and results in Section V, followed by a review of related works in Section VI. Finally, we conclude our proposed *MORE* approach in Section VII.

II. PRELIMINARIES

A. Quantum state and visualization

A qubit (short for a quantum bit) is the information carrier in the quantum computing/communication channel [40]. A qubit is defined as a two-dimensional Hilbert space with two orthonormal bases $|0\rangle$ and $|1\rangle$, which are known as computational bases in two-level quantum computing. These computational bases are usually represented as vectors $|0\rangle = [1, 0]^T$ and $|1\rangle = [0, 1]^T$. Due to the unique qubit characteristic of *superposition*, the state of a qubit can be represented as the sum of two computational bases weighted by (complex) amplitudes:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (1)$$

where α and $\beta \in \mathbb{C}$, and $|\alpha|^2 + |\beta|^2 = 1$. $|\alpha|^2$ and $|\beta|^2$ are the probability of obtaining states $|0\rangle$ and $|1\rangle$ after multiple measurements, respectively.

The Bloch sphere is a valuable tool for visualizing the state of a single qubit, as shown in Fig. 2. It encompasses all possible states of a qubit, making it an excellent representation of a two-dimensional Hilbert space. In this article, we will use the Bloch sphere to illustrate our proposed method clearly. Every pure state of a qubit can be mapped to a distinct point on the surface of the Bloch sphere, whereas mixed states correspond to points within the sphere. The state of a qubit on the Bloch sphere can be described with two real parameters, θ and ϕ ,

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \quad (2)$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. I.e., $\theta = 0$ for $|0\rangle$ and $\theta = \pi$ for $|1\rangle$, and global phase ϕ can be any value.

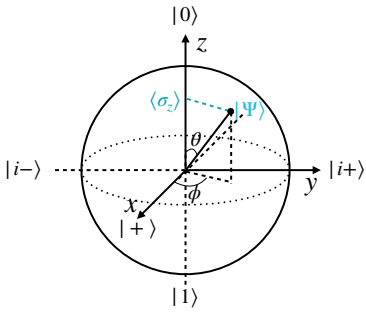


Fig. 2. Bloch sphere and z-measurement

B. Quantum measurement

Quantum measurement is the retrieval of the numerical information stored in a qubit. A measurement result is +1 for state $|0\rangle$ and -1 for state $|1\rangle$ according to a specified probability distribution associated with the quantum state. Therefore, numerous measurements are required to determine the exact quantum state. The final result of the quantum measurement is the expected value of all outcomes.

Observables are used to understand the properties of a quantum system and can be measured. Mathematically, observables are formulated as Hermitian operators that map Hilbert space onto themselves. For a valid observable, its eigenvalues are real numbers and can be the outcomes of measurement. Moreover, observables can form an orthonormal basis of the target Hilbert space, which will be the state of the quantum system after measurement. The observables considered in this article are Pauli matrices:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (3)$$

These Pauli matrices span a complex two-dimensional Hilbert space (a qubit). The projection measurement is to extract quantum information by operating on the interested observable and the density matrix of the target quantum state

$$\langle \sigma \rangle = \text{Tr}(\sigma |\psi\rangle\langle\psi|) \quad (4)$$

where $|\psi\rangle\langle\psi|$ generates the density matrix. For the general z-measurement, the state vector is projected onto the z-axis of the Bloch sphere, and the corresponding value on the z-axis is the expectation of the measurement results, as shown in Fig. 2.

C. Variational quantum algorithm

The variational quantum algorithm (VQA) is the standard approach to performing QNN. It processes prepared quantum information by applying a series of parametric quantum gates, ultimately producing an output through measurement. As an example, a binary quantum classifier is implemented using VQA in [13]. It has an $(n+1)$ -qubit circuit, where the first n qubits are prepared using an encoding method (such as angle, basis, or amplitude encoding) to represent specific information. The final qubit acted as a readout, generating the output through measurements. These qubits then pass a sequence of

quantum gates with trainable parameters $U(\theta) = \prod_{l=1}^N U_l(\theta_l)$, where θ is a set of parameters. A measurement outcome of 1 corresponds to one class, while a result of -1 corresponds to the other class. VQA uses a hybrid quantum-classical procedure to iteratively optimize the trainable parameters. The popular optimization approach includes gradient descent [41], parameter shift [42], and gradient-free techniques, such as COBYLA. All of the methods take the training data as input and evaluate the model performance by comparing the generated and correct labels. Based on this evaluation, the methods update the model parameters for the next round, repeating the process until the model converges and achieves the desired performance. The hybrid method performs the evaluation and parameter selection on a classical machine, while the model inference is carried out on a quantum machine.

III. QUANTUM STATE TOMOGRAPHY

Quantum state tomography (QST) is a technique to reconstruct an unknown quantum state using its measurement results [43]. The measured observables must form a basis in the Hilbert space so that all state information can be recorded and used to recover the state. In the present age of NISQ, when the number of qubits is limited, QST is an essential method for retrieving the complete information stored in a quantum system. Nevertheless, as the number of qubits grows, the number of measurements needed and the complexity of state reconstruction increase exponentially. Hence, for the sake of simplicity, we consider the reconstruction of a single qubit using three observables to restore its state in a two-dimensional Hilbert space in this work.

Any arbitrary density matrix of a 1-qubit state can be expressed as a linear combination of Pauli matrices (basis of two-dimensional Hilbert space) as

$$\begin{aligned} \rho &= \frac{1}{2}(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z) \\ &= \frac{1}{2} \begin{bmatrix} r_z + 1 & r_x - r_y \\ r_x + r_y & -r_z + 1 \end{bmatrix} \end{aligned} \quad (5)$$

where r denotes real number and $r_x^2 + r_y^2 + r_z^2 = 1$. For the typical qubit measurement using observable σ_z , as used by most quantum applications, the expectation of the measurement result is

$$\langle \sigma_z \rangle = \text{Tr}\left(\frac{1}{2}(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z)\sigma_z\right) = r_z \quad (6)$$

Eq. 6 demonstrates that the expectation of measurement results is directly related to the density matrix ρ of interest. In this case, however, only diagonal entries of ρ can be retrieved, while some useful information remains untouched. Hence, in order to rebuild the density matrix ρ completely, the measurements on observables σ_x and σ_y are necessary to obtain r_x and r_y .

From the perspective of the geometric representation, the expectation values $\langle \sigma_x \rangle$, $\langle \sigma_y \rangle$ and $\langle \sigma_z \rangle$ are exactly the projections of the state vector on the x-, y- and z-axis of the Bloch sphere, respectively. Therefore, after getting sufficient

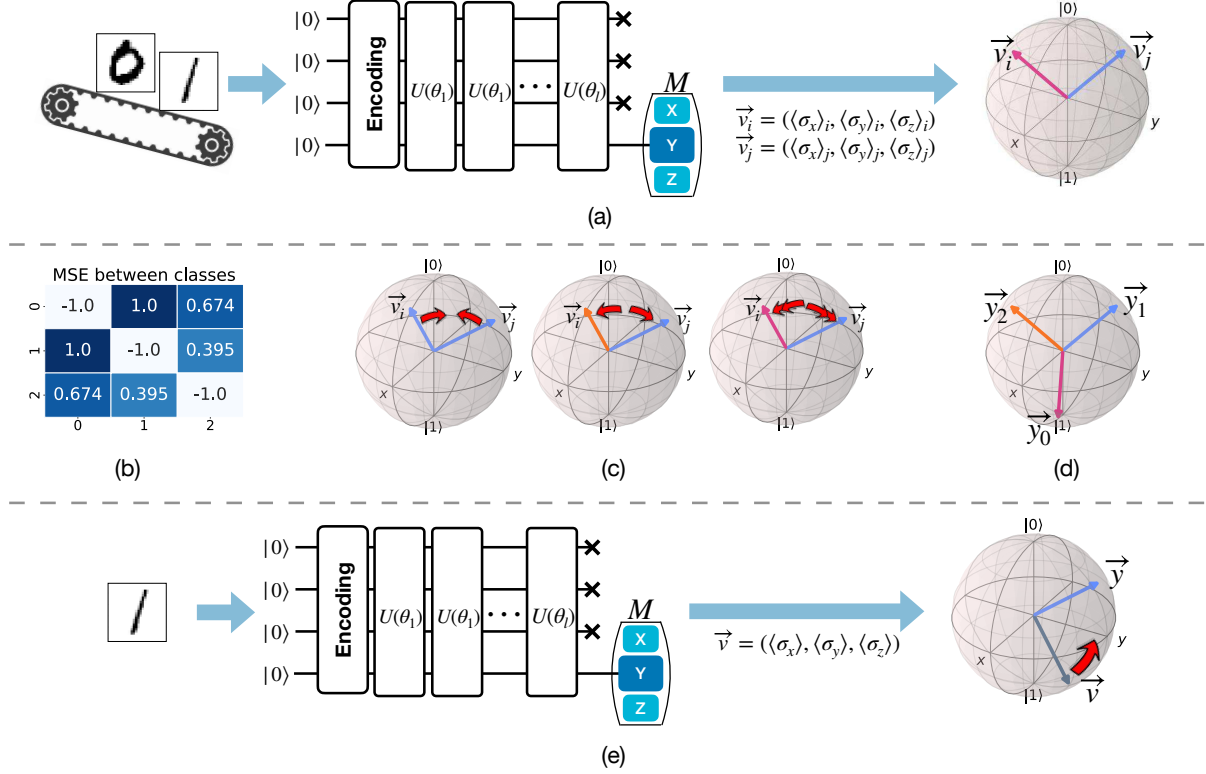


Fig. 3. MORE overview. Step 1: variational quantum clustering (a)-(d). (a) The QNN takes a pair of training instances as input and reconstructs their single-qubit readout states using x-, y-, and z-measurements. (c) QNN's parameters are tuned to adjust the distance between readout states depending on (b) the interclass correlations represented by MSE. (d) After iterative training, the distribution of readout states in Hilbert space corresponds to that of classes in feature space. The centroid of readout states within the same class is the quantum label for this class. Step 2: quantum label-based supervised training (e). The readout state tends to approach its corresponding quantum label during the training process. During inference, the predicted label of a test instance is the quantum label closest to its readout state.

measurement results, we can restore the quantum state in the Bloch sphere for an intuitive interpretation. In other words, the interested 1-qubit state is a state vector with an unknown direction in the Bloch sphere. By projecting the vector on the three axes, we can figure out the direction of the state vector, which contains all information about the state.

IV. MEASUREMENT AND CORRELATION-BASED QNN MULTI-CLASSIFIER

In this section, we present MORE, an efficient quantum multi-classifier that employs a simple QNN with a single readout qubit. We use QST to reconstruct the state of the readout qubit to retrieve the entire information from it. This allows us to fully leverage information of a single qubit to address complex multi-classification tasks via two steps. First, we employ the variational quantum clustering method to convert the classical labels of the dataset into 1-qubit states, taking into account interclass correlations. Next, we proceed to train the QNN using a quantum label-based supervised learning approach, where the QNN learns to map an input to its corresponding quantum label. After completing the training process, the model can accurately categorize new data by comparing the readout state of the data with quantum labels

and assigning it to the appropriate class based on the shortest distance.

A. Overview

The overview of MORE is presented in Fig. 3. The goal is to solve a multi-classification task that assigns appropriate labels to classical input data. A training dataset consisting of K subsets is provided, $D = \{D_1, D_2, \dots, D_K\}$. The subset D_k is a collection of training data in class k , with each training data instance being a pair of data point x and its correct classical label y . MORE addresses this problem by utilizing a QNN to learn the pattern that characterizes the mapping between the input data point and its correct label from the training dataset. The learning of MORE is a two-step procedure. First, the classical labels y are converted into quantum labels $\vec{y}$ in the two-dimensional Hilbert space. This step employs the quantum clustering method while considering the inherent correlation between classes. (Fig. 3(a-d) and Section IV-B). Then, quantum label-based supervised learning is undertaken to fine-tune the parameters (Fig. 3(e) and Section IV-C).

The ansatz of MORE is shown in Fig 3(a). Its circuit includes three layers: (1) Encoding layer: the encoding layer quantizes the classical input data by preparing the quantum

states based on classical data values. MORE is compatible with any general encoding scheme, such as angle and amplitude encoding. (2) Variational data processing layer $U(\theta)$: this layer transforms the prepared state using its parameterized gates. The parameter set θ of the layer is iteratively adjusted based on MORE's performance using a hybrid classical-quantum method, as introduced in section II. The optimal parameters should enable MORE to assign a label y to unseen input x with high accuracy. (3) Measurement layer M : One qubit is chosen as the readout in this layer, and its observable σ_x , σ_y , and σ_z are measured to produce a three-dimensional vector (reconstructed state vector $\vec{v}_i$ and $\vec{v}_j$).

As the quantum measurement operator in superconducting hardware implementation can only measure the state on the z-axis, MORE requires two additional measurement units to measure the observables σ_x and σ_y , as shown in Fig. 4. The unit for observable σ_x inserts an H gate before the measurement operator, which can be conceptualized as rotating the x-axis of the Bloch sphere to the location of the z-axis first and then projecting the state vector to the new z-axis. Similarly, for observable σ_y the readout qubit passes an $S^\dagger$ gate and an H gate, followed by the measurement operator in the measurement unit. In the measurement layer of the QNN, these three measurement units are applied in succession to construct the state vector $\vec{v} = (\langle\sigma_x\rangle, \langle\sigma_y\rangle, \langle\sigma_z\rangle)$.

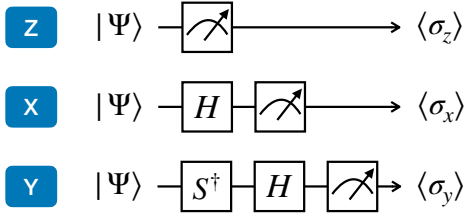


Fig. 4. Measurement unit of MORE

B. Class correlation-based variational quantum clustering

To perform multi-classification on the 1-qubit state vectors, it is necessary to determine the distribution of classes in a two-dimensional Hilbert space first. This involves converting classical labels to quantum labels. To accomplish this, we use the *variational quantum clustering* method while investigating the correlation between classes. Deep clustering is a technique in classical machine learning that concurrently learns the parameters of neural networks and the cluster assignments for processed inputs [44]. Here, we borrow this idea and apply it to variational quantum clustering. Specifically, we train the parameters of the QNN to map readout state vectors from the same class to a particular quantum state.

Class correlation. First of all, we calculate the correlation between classes by measuring the difference between the data instances belonging to different class collections. To be specific, we first pre-process the training data using Principal Component Analysis (PCA) to extract critical features and downscale the data to a size that is suitable for our quantum

circuit. Then, the average pattern set $D_{pattern}$ is created by averaging the data instances of the class dataset D_k ,

$$D_{pattern} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_K\} \quad (7)$$

where $\bar{x}_k = \frac{1}{|D_k|} \sum_{i=1}^{|D_k|} x_k^i$, $k \in [1, K]$. The correlation between the average pattern of classes is then recorded in a $K \times K$ array, known as scaler array S . For example, we use Mean square error (MSE) to quantify the difference between two average patterns. As a result, the entry (i, j) of the scaler array S can be calculated by

$$S_{i,j} = MSE(\bar{x}_i, \bar{x}_j) \quad (8)$$

The larger the MSE value, the lower the correlation. The original diagonal entries of S are 0s, indicating that the correlation between two identical patterns is maximal. We modify these 0s to negative values, which will be explained and utilized later in this section. Fig. 3(b) shows an example of the scaler array S that stores the correlations between classes '0', '1', and '2' from the MNIST handwritten digit dataset.

Clustering dataset. We then prepare the clustering dataset by pairing the instances in the given dataset D . The clustering dataset consisting of data pairs: $D_{cluster} = \{(x_i, y_i), (x_j, y_j)\}_{t=1}^N$ where $i, j \in [1, K]$ and N is a small number to ensure the model can fast converge. In our evaluation, we set it to $\binom{5K}{2}$. Each data pair is independently sampled from D across all classes, i.e., (x_i, y_i) and (x_j, y_j) are two sampled instances from class i and j .

Clustering loss. In a training step, the QNN $U(\theta)$ sequentially takes as input of a data pair x_i and x_j and generates two three-dimensional state vectors $\vec{v}_i = (\langle\sigma_x\rangle_i, \langle\sigma_y\rangle_i, \langle\sigma_z\rangle_i)$ and $\vec{v}_j = (\langle\sigma_x\rangle_j, \langle\sigma_y\rangle_j, \langle\sigma_z\rangle_j)$. We express the loss function as

$$\mathcal{L}_{clt} = -S_{i,j} \times Dist(\vec{v}_i, \vec{v}_j) \quad (9)$$

where

$$Dist(\vec{v}_i, \vec{v}_j) = 1 - \frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|} \quad (10)$$

The function $Dist(\vec{v}_i, \vec{v}_j)$ computes the Cosine distance between two vectors $\vec{v}_i$ and $\vec{v}_j$, and the coefficient $S_{i,j}$ serves as the scaler of cosine distance. We assume that the distribution of state vectors in a two-dimensional Hilbert space is relevant to the distribution of data in feature space. Therefore, the clustering step aims to bring the generated state vectors $\vec{v}_i$ and $\vec{v}_j$ closer together if x_i and x_j belong to the same class, and farther apart otherwise. In addition, the distance between $\vec{v}_i$ and $\vec{v}_j$ will be larger if their class correlation is significantly smaller. Specifically, as illustrated in Fig. 3(c), if x_i and x_j from the same class, the negative scaler value $S_{i,j}$ where $i = j$, will minimize the cosine distance between $\vec{v}_i$ and $\vec{v}_j$. Otherwise, the distance between $\vec{v}_i$ and $\vec{v}_j$ will be maximized, and the final distance between them depends on the value of $S_{i,j}$. The higher the correlation, the closer they will be.

Quantum labels. After the training of quantum clustering, the state vectors of data instances belonging to the same class are oriented in similar directions (a cluster). The centroid of these directions is recorded as the quantum label of the class.

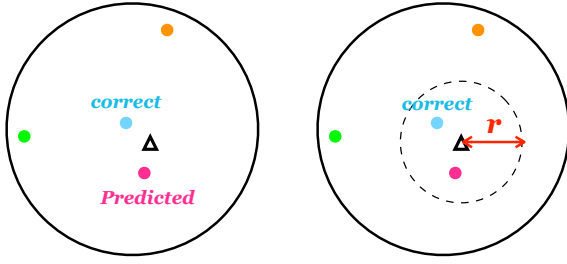


Fig. 5. Curse of density (left) and loss adjuster (right)

So far, we have mapped the classical labels y into quantum labels $\vec{y}$, which are distributed in the Hilbert space based on the correlation between classes, as shown in Fig. 3(d). Then we substitute the classical labels y in the training dataset D with quantum labels $\vec{y}$, and now we are ready to proceed with quantum label-based quantum supervised learning.

C. Quantum label-based supervised learning

After the clustering step, we initially train the parameters θ of QNN $U(\theta)$ to determine the quantum labels for classes. However, as the $U(\theta)$ is trained on a small subset of the training data, it may not be optimal for classification. Thus, to enhance the performance of $U(\theta)$, we fine-tune the parameters by conducting quantum label-based supervised learning. We use the dataset containing the training data points x and their quantum labels $\vec{y}$ to train $U(\theta)$. The goal is to map the readout state vector of inputs to be as close as possible to their corresponding quantum labels, as shown in Fig. 3 (e). In this step, the $U(\theta)$ takes one training instance as input and generates the 1-qubit state vector $\vec{v}$. Then, we compare the generated vector with its corresponding quantum label $\vec{y}$, and express the loss function of quantum label-based supervised learning as

$$\mathcal{L}_{sup} = \text{Dist}(\vec{v}, \vec{y}) \quad (11)$$

The training procedure aims to minimize the value of $\mathcal{L}_{sup}$. For inference, the class of the unseen data is determined by

$$y = \text{argmin}_k \{ \text{Dist}(\vec{v}, \vec{y}_k) \} \quad (12)$$

where $\vec{y}_k$ is the quantum label for class $k \in [1, K]$.

However, misclassifications are more likely to occur when there are a large number of classes and their quantum labels are closely distributed in the Hilbert space, which is known as the curse of density (see Fig. 5 left). The circle is a side view of the Bloch sphere, and the colored dots are quantum labels. The triangle represents the readout state of an input. It may be classified into a nearby but incorrect class (represented by the pink dot) due to the small distance between the readout state and the pink quantum label. To improve the accuracy of the classifier in scenarios with crowded quantum labels, we introduce a loss adjuster $\mathcal{R}$.

In each training step, if the distance between the readout state and its correct quantum label is less than a threshold r , we also consider the other quantum labels whose distance

from the readout state is less than the threshold, as shown in Fig. 5 right. Specifically, when $\text{Dist}(\vec{y}, \vec{v}) \leq r$, we define the loss regulator as

$$\mathcal{R} = \sum_{k \in K'} \text{Dist}(\vec{y}_k, \vec{v}) \quad (13)$$

where K' is a subset of classes whose quantum label $\vec{y}_k$ satisfies $\text{Dist}(\vec{y}_k, \vec{v}) \leq r$.

We then assign a weight factor $w \in [0, 1]$ to the supervised loss term. Hence, the overall objective function of the quantum labels-based training process is

$$\mathcal{L}_{sup+\mathcal{R}} = w\mathcal{L}_{sup} - (1 - w)\mathcal{R}. \quad (14)$$

By minimizing the value of $\mathcal{L}_{sup+\mathcal{R}}$, this objective function tends to bring the readout state closer to its correct quantum label while moving farther away from the incorrect quantum labels surrounding it, resulting in the readout state being closest to its correct label among crowded quantum labels. The selection strategy for the values of r and w will be discussed in Sec. V-C.

V. EVALUATION

We conducted empirical studies in various scenarios to evaluate the performance of the proposed quantum multi-classifier MORE. To implement MORE and related baselines, we used Python 3.8 and the IBM Qiskit package [45] to simulate quantum systems both with and without noise. For simulating noisy systems, we employed several noisy backends, including FakeAuckland, FakeAthensV2, and FakeBelemV2. The source code used to generate the experiment results is available in github.com/Jindi0/MORE.

Dataset: (1) We utilize the **MNIST** dataset [46] to conduct experiments in the noiseless quantum system. The MNIST is a widely used benchmark for image classification tasks, consisting of ten classes of hand-written digits ranging from 0 to 9. We randomly select 1,000 training and 200 test images from each class and reduce their dimensions to 8 using PCA. (2) Additionally, we evaluate MORE using the **Iris** dataset [47] in simulated noisy quantum systems. The Iris dataset contains 150 instances classified into three distinct classes, each consisting of 50 instances with 4 pixels. We use 70% of the instances (105 instances) for training, while the remaining 30% (45 instances) are reserved for testing purposes.

Model implementation: We develop the QNN classifiers based on Qiskit `NeuralNetworkClassifier` class, and use the optimizer `COBYLA` to update trainable parameters. MORE employs an 8-qubit QNN for the MNIST dataset and a 4-qubit QNN for the Iris dataset, with 91 and 39 trainable parameters, respectively. We build QNNs based on the design principles for quantum convolutional neural networks proposed in [30]. For the readout, we choose the last active qubits at the end of the circuit and measure it with σ_x , σ_y , and σ_z observables. To efficiently convert classical labels to quantum labels during the clustering step, we randomly select five instances from each class and form pairs from the resulting dataset. The clustering dataset consists of $\binom{5K}{2}$ pairs in total,

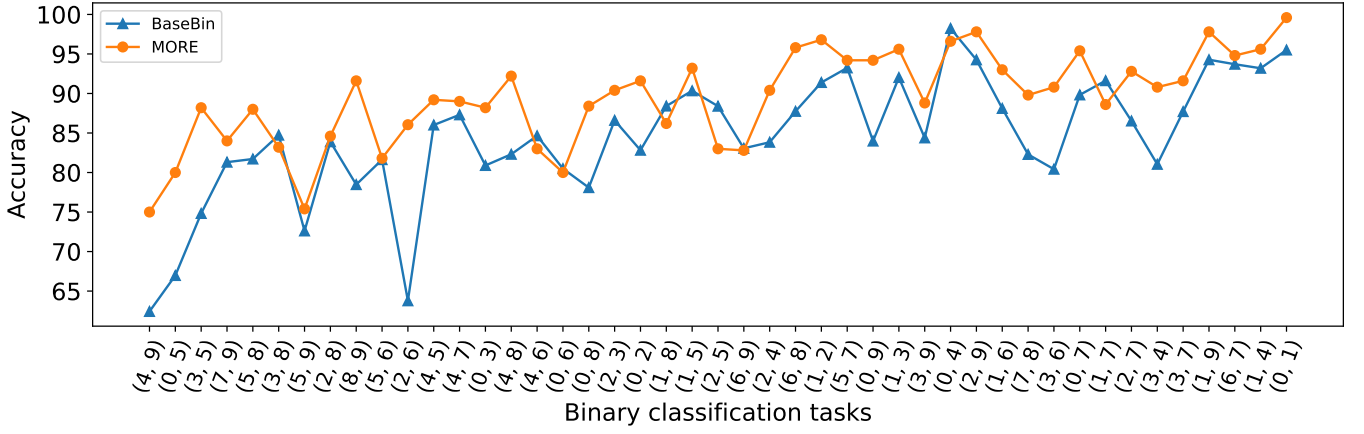


Fig. 6. Accuracy of binary classification tasks

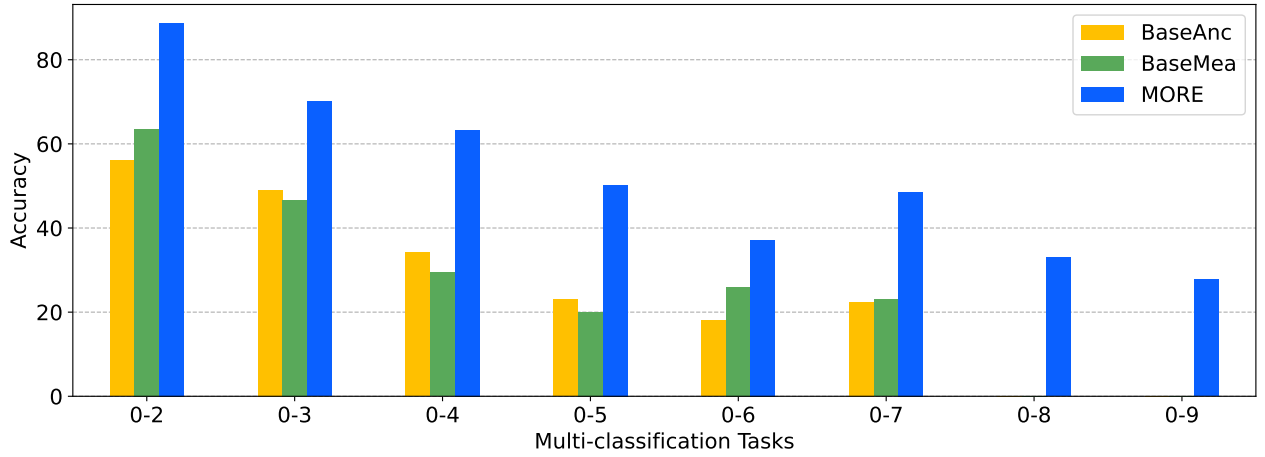


Fig. 7. Accuracy of multi-classification tasks

where K is the number of classes. We use MSE to calculate the interclass correlations. And to make the quantum labels spread out as much as possible, the MSE between different classes is normalized to the interval $[0.5, 1]$. During supervised learning, then, the entire training dataset is used.

A. Accuracy

We evaluate the accuracy of the proposed approach **MORE** using the MNIST dataset on noise-free quantum systems. We conduct nine classification problems ranging from binary to 10-class.

Our evaluation starts by investigating the effect of using multiple observables of a readout qubit on binary classification tasks. To do so, we compare our results to **BaseBin**, which serves as the baseline in this experiment. The comparison is illustrated in Fig. 6. Both MORE and BaseBin use the same ansatz (QCNN with 91 trainable parameters) for binary classifications, but the readout qubit of BaseBin is measured only with observable σ_z . The expected value of its measurement results is associated with two distinct classes: a positive expected value represents one class, and a negative expected

value represents the other. We conduct 45 binary classifications on MNIST using MORE and BaseBin, respectively. The class pairs for classification are listed on the x-axis of Fig. 6 and are sorted in descending order of interclass correlations. E.g., the training data of classes ‘4’ and ‘9’ have the highest similarity, while those of classes ‘0’ and ‘1’ have the lowest. The results indicate that MORE outperforms BaseBin in 37 out of 45 tasks and achieves comparable accuracy in the remaining tasks. MORE improves accuracy by up to 22.28% and by an average of 4.9%. Furthermore, the performance of MORE demonstrates greater stability across tasks of varying difficulty than BaseBin. As observed, a roughly inverse relationship exists between accuracy and interclass correlation for both MORE and BaseBin, i.e., as class correlation increases, classification becomes more challenging. Consequently, both MORE and BaseBin exhibit the highest accuracy on (0, 1)-classification and the lowest accuracy on (4, 9)-classification. Nonetheless, the variance of MORE’s accuracy over 45 tasks is only 34.01, while that of BaseBin is 59.41, indicating that MORE is more stable than BaseBin. It shows the advantage of MORE, which

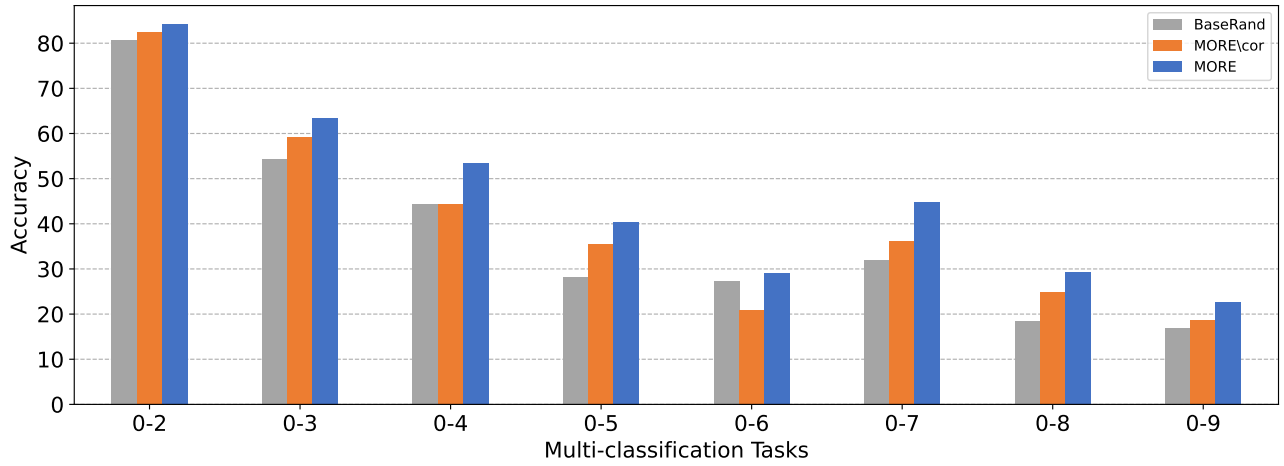


Fig. 8. Test accuracy of multi-classifications over various quantum label selection strategies

has three observables, in terms of accuracy and stability.

Next, we evaluate MORE on multi-classification tasks using the MNIST dataset and summarize the results in Fig. 7. We conduct eight multi-classification tasks, categorizing handwritten digits into three to ten classes. Note that the ansatzes used in multi-classifications are *identical* to the binary classification ansatzes. The accuracy of MOREs is compared to that of **BaseAnc** and **BaseMea**, which serve as the baselines in this experiment. BaseAnc and BaseMea are both variational quantum multi-classifiers. BaseAnc utilizes eight qubits for data processing and n ancilla qubits as the readout for n -class classifications. And BaseMea employs a total of eight qubits and measures a subset of qubits at the end of the circuit to produce a result. In a multi-classification task with n classes, n qubits are measured. They encode class labels as one-hot vectors and measure the readout qubits on the z-basis. So BaseAnc and BaseMea can only support the classifications involving up to eight classes, due to the limitation in dimensions and the number of qubits, respectively. MORE, however, is capable of conducting 10-class classification regardless of the number of qubits, and we believe it still qualifies if the dataset contains more classes. Fig. 7 indicates that MORE outperforms BaseAnc and BaseMea across the board. The reason is that both BaseAnc and BaseMea require simultaneous control of several qubits to represent the class label, which can be challenging, particularly for an unstable quantum system. In contrast, MORE only utilizes a single qubit to record the outcome, thus reducing the number of factors contributing to instability.

Overall, our proposed MORE approach, which uses a simple circuit with only one readout qubit, can beat other general approaches and achieve the desired performance. In the following subsections, we will analyze the impact of the MORE's components and access MORE in noisy quantum systems.

B. Quantum label

We now examine the impact of selecting quantum labels for classes. To do so, we compare the following approaches:

- **BaseRand**: A baseline method that randomly assigns quantum states to classical labels.
- **MORE\cor**: A variant of the MORE approach that does not consider interclass correlation during the variational quantum clustering step, i.e., all diagonal entries in the scaler array S are -1s and all other entries are 1s.
- **MORE**: The vanilla MORE method employs interclass correlation for determining quantum labels.

Then, based on the selected quantum labels, each of the three approaches is followed by quantum label-based supervised learning using objective function Eq. 11 without the loss regulator $\mathcal{R}$. The test accuracy of the above approaches is summarized in Fig. 8. Across eight multi-classification tasks, MORE outperforms other methods in all cases. BaseRand and MORE\cor have comparable performance, all of which are inferior to MORE. Therefore, we conclude that it is beneficial to take interclass correlations into account when deciding the quantum labels for classical labels. A possible reason is that the data distribution in the Hilbert space is strongly related to that in the classical feature space. As an example, Fig 3 (d) shows the distribution of quantum labels for class '0', '1', and '2' of MNIST, according to their relationship as listed in Fig 3 (b). Classes '1' and '2' have the smallest MSE value (strongest correlation), so their quantum labels are closer to each other compared to other class pairs. Similarly, the readout states of the instance from classes '1' and '2' are supposed to be closer than those of other classes. As a result, assigning quantum labels based on class correlation can capture the training data pattern to some extent, which is advantageous for enhancing model quality, as shown in Fig. 8. Nonetheless, the figure reveals that the accuracy of quantum classifiers decreases as the number of classes increases. This trend is also observed in quantum classifiers that employ alternative implementation strategies. Thus, we introduce a loss adjuster to alleviate this issue, as shown in Eq. 13, and analyze its impact in the next subsection.

TABLE I
PARAMETERS AND RESULTS FOR LOSS ADJUSTER

Task	Min. label distance	r	w	More\R acc.	MORE acc.
0-2	1.302	1.5	0.1	84.13%	88.7%
0-3	0.713	1.0	0.2	63.45%	70.1%
0-4	0.57	0.8	0.1	53.5%	63.3%
0-5	0.077	0.2	0.4	40.48%	50.2%
0-6	0.212	0.4	0.5	29.06%	37.2%
0-7	0.067	0.2	0.5	44.8%	48.6%
0-8	0.099	0.15	0.2	29.4%	33%
0-9	0.086	0.15	0.2	22.6%	27.8%

C. Loss adjuster

Table I summarizes the parameters and results for the loss adjuster over the multi-classification tasks. The accuracy of **MORE** (with $\mathcal{R}$) and **MORE\R** (without $\mathcal{R}$) are listed in the last two columns. The comparison indicates that **MORE**, which uses the loss function (Eq. 14) with $\mathcal{R}$ during supervised learning, improves the accuracy across all tasks. This improvement is caused by the corrected misclassification between the classes whose quantum labels in the Hilbert space are too near together. In Eq. 14, two hyper-parameters need to be determined by users: the threshold r and the weight w . The threshold r is empirically determined based on distances between quantum labels. In our experiments, the Cosine distance between quantum labels is calculated. Usually, r is specified to be slightly larger than the shortest distance, as shown in the 2nd and 3rd columns of table I. E.g., the smallest Cosine distance between quantum labels in the 0-9 task, a 10-class classification, is 0.086 (about 24 degrees), so we set the r to be 0.15 (about 32 degrees around the target quantum label). Moreover, we test **MORE** with a w range from 0.1 to 1.0 and identify the optimal value of w for each task. We report these values in the 4th column of the table, although we found accuracy improvements across all values of w in each task. The w values demonstrate that $\mathcal{R}$ has varying importance across different tasks, but its contribution is not greater than half in any task. Hence, we recommend that users search for an appropriate value of w within the range of 0.1 to 0.5.

D. Noisy quantum system

Table II summarizes the test accuracy of **MORE** using the noisy quantum backends, including FakeAuckland, FakeAthensV2, and FakeBelemV2 on the Iris dataset. The hybrid quantum-classical training method of QNN typically requires a long time to execute on a quantum machine, but the IBM cloud quantum computing platform imposes time limits for tasks. In this experiment, we therefore utilize simulators of noisy quantum computing systems. The noisy backends used in this experiment have the noise model collected from real quantum machines, which includes T1 and T2 time, 1-qubit and 2-qubit gate errors, and measurement errors. So these

TABLE II
TEST ACCURACY OF MORE ON IRIS DATASET WITH NOISY BACKENDS

# Shots	FakeAuckland	FakeAthensV2	FakeBelemV2
1k	95.56	95.56	93.33
2k	97.78	93.33	95.56
3k	97.78	93.33	95.56
4k	97.78	93.33	95.56
5k	93.33	93.33	95.56
6k	95.56	95.56	95.56

simulated backends provide us with a practical and reasonable way to evaluate the performance of **MORE** in noisy quantum systems. To process the Iris instances of size 4, we construct 4-qubit QNNs with 39 trainable parameters. We update the QNNs for 300 steps using the COBYLA optimizer. We varied the number of shots for the quantum program from 1,000 to 6,000 on the backends to examine the impact of quantum computation costs.

The MOREs reach their highest accuracies of 97.78 (44/45), 95.56 (43/45), and 95.56 (43/45) on the FakeAuckland, FakeAthensV2, and FakeBelemV2 backends, respectively. The noisy MOREs achieve comparable performance to the noise-free version, with an accuracy of 97.78%. Surprisingly, we found that the accuracy was not directly proportional to the number of shots, suggesting that the quantum computation cost required for this task is not significant. However, we acknowledge that the number of shots required may increase with an increase in the number of classes. Nonetheless, the cost is still acceptable only if the distribution of learned quantum labels is scattered in the Hilbert space. Furthermore, while it is true that accurate measurement results are necessary for the **MORE** approach, we want to highlight that using a simple quantum circuit (with fewer qubits, gates, and shallow depth) can lead to less error compared to other methods. This makes the **MORE** approach feasible and promising to implement on NISQ machines.

VI. RELATED WORK

In recent years, numerous attempts have been made to explore the capabilities of QNNs using various techniques. In this section, we will discuss some notable works that focus primarily on the classification task, which is the central theme of this article.

First of all, there are several suggested concepts for the design of QNN-based classifiers [13], [30], [31], [35]. [13] and [30] concentrate on the ansatz design of QNNs. They are the fundamentals of current popular QNNs that use parametric unitary transformations to process data efficiently. However, scalability is a crucial concern for QNNs, particularly on NISQ devices, where the limited number of qubits restricts the quantum circuit's size and performance. Quconvolutional NN [35] employs the quantum circuit of modest size as kernels of quantum convolution NN, rather than building the entire QNN. And SQNN [31] scales up QNNs by constructing large-

scale QNNs modularly and leveraging quantum computation resources from multiple quantum machines. All of these studies focus on binary classification tasks and demonstrate the advantages of QNNs. In addition to model scalability, the ability to scale the problem size is essential for improving the capacity of QNNs. In this article, we propose an approach that aims to expand the problem size that a QNN can handle without having to scale up its quantum circuit.

Many prior efforts have attempted to enhance the problem size that QNN classifiers can address [33], [36], [37], [48]–[53]. A QNN classifier with four ancilla qubits for 4-class classification is built in [48]. The capability of the solution is restricted by the quantum hardware, as larger problem sizes require additional ancilla qubits, which are not currently available on NISQ machines. [33], [53] also address multi-classification problems by using ancilla qubits. An alternative solution is QuClassi proposed in [39]. It breaks down a multi-classification into several binary classifications. For each binary classification, the cost function is constructed based on quantum state fidelity by using SWAP, and only one ancilla qubit is needed for measurement to get the readout. The measurement results of all classes are softmaxed to obtain the final classification result. Despite the quantum computing resources being limited in this method, the training and inference processes are complicated and time-consuming. Similarly, a fidelity-based QNN classifier is proposed in [37], but its QNN ansatz can only solve binary classification tasks. Moreover, VSQ [49] slides a predefined partial observable “xx” across the qubits to get the classical shadows of the quantum state prepared by a QNN, and then feeds the local shadows to a classical fully-connected layer for decision making. Yet, the computationally demanding classical fully-connected layer may counteract the quantum speedup. A unified quantum classifier is presented in [50]. This approach investigates the connection between class labels and quantum states. In our study, we further investigate the distribution of quantum labels in Hilbert space based on interclass correlations and demonstrate its significance by highlighting the superior performance of our approach. Additionally, we introduce a loss adjuster during supervised learning to ensure optimal performance when dealing with multiple classes in the task at hand.

VII. CONCLUSION

We propose MORE, a QNN-based multi-classifier that maximizes quantum resource efficiency. By fully leveraging quantum information, MORE employs a simple ansatz as the binary classifier with only one readout qubit to solve multi-classification problems. To achieve this, MORE first converts classical labels into corresponding quantum labels using variational quantum clustering. This process considers interclass correlations, allowing the learned quantum labels to capture intrinsic patterns in the classical feature space of the training data. Quantum supervised learning is then performed based on the quantum labels to efficiently train the model. Furthermore, we introduce a loss adjuster to enhance the model’s quality by making it more sensitive to the labels

that can result in misclassifications during training. Our comprehensive evaluations demonstrate that MORE outperforms general quantum multi-classifiers and is capable of effective operation in noisy quantum systems due to its reduced error sources. This suggests that fully utilizing quantum information offers a promising approach for scaling up the problem sizes that can be addressed during the current NISQ era.

ACKNOWLEDGMENT

The authors would like to thank all the reviewers for their helpful comments. This project was supported in part by US National Science Foundation grant CNS-1816399. This work was also supported in part by the Commonwealth Cyber Initiative, an investment in the advancement of cyber R&D, innovation and workforce development. For more information about CCI, visit cyberinitiative.org.

REFERENCES

- [1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.
- [2] H. Collins and C. Nay, “Ibm unveils 400 qubit-plus quantum processor and next-generation ibm quantum system two,” <https://newsroom.ibm.com/2022-11-09-IBM-Unveils-400-Qubit-Plus-Quantum-Processor-and-Next-Generation-IBM-Quantum-System-Two>, 2022.
- [3] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, “Introduction to quantum noise, measurement, and amplification,” *Reviews of Modern Physics*, vol. 82, no. 2, p. 1155, 2010.
- [4] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.
- [5] M. Schuld, I. Sinayskiy, and F. Petruccione, “An introduction to quantum machine learning,” *Contemporary Physics*, vol. 56, no. 2, pp. 172–185, 2015.
- [6] E. Aïmeur, G. Brassard, and S. Gambs, “Machine learning in a quantum world,” in *Conference of the Canadian Society for Computational Studies of Intelligence*. Springer, 2006, pp. 431–442.
- [7] Y. Dang, N. Jiang, H. Hu, Z. Ji, and W. Zhang, “Image classification based on quantum k-nearest-neighbor algorithm,” *Quantum Information Processing*, vol. 17, no. 9, pp. 1–18, 2018.
- [8] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Physical review letters*, vol. 113, no. 13, p. 130503, 2014.
- [9] Z. Li, X. Liu, N. Xu, and J. Du, “Experimental realization of a quantum support vector machine,” *Physical review letters*, vol. 114, no. 14, p. 140504, 2015.
- [10] R. Chatterjee and T. Yu, “Generalized coherent states, reproducing kernels, and quantum support vector machines,” *arXiv preprint arXiv:1612.03713*, 2016.
- [11] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, “Quantum cluster theories,” *Reviews of Modern Physics*, vol. 77, no. 3, p. 1027, 2005.
- [12] E. Aïmeur, G. Brassard, and S. Gambs, “Quantum clustering algorithms,” in *Proceedings of the 24th international conference on machine learning*, 2007, pp. 1–8.
- [13] E. Farhi and H. Neven, “Classification with quantum neural networks on near term processors,” *arXiv preprint arXiv:1802.06002*, 2018.
- [14] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, “Training deep quantum neural networks,” *Nature communications*, vol. 11, no. 1, pp. 1–6, 2020.
- [15] S. Jeswal and S. Chakraverty, “Recent developments and applications in quantum neural network: a review,” *Archives of Computational Methods in Engineering*, vol. 26, no. 4, pp. 793–807, 2019.
- [16] M. A. Shahin, M. B. Jaksa, and H. R. Maier, “Artificial neural network applications in geotechnical engineering,” *Australian geomechanics*, vol. 36, no. 1, pp. 49–62, 2001.

- [17] J. Fu, R. Yang, X. Li, X. Sun, Y. Li, Z. Liu, Y. Zhang, and B. Sunden, "Application of artificial neural network to forecast engine performance and emissions of a spark ignition engine," *Applied Thermal Engineering*, vol. 201, p. 117749, 2022.
- [18] A. Lakhan, Q.-U.-A. Mastoi, M. Elhoseny, M. S. Memon, and M. A. Mohammed, "Deep neural network-based application partitioning and scheduling for hospitals and medical enterprises using iot assisted mobile fog cloud," *Enterprise Information Systems*, vol. 16, no. 7, p. 1883122, 2022.
- [19] E. Elahi, Z. Zhang, Z. Khalid, and H. Xu, "Application of an artificial neural network to optimise energy inputs: An energy-and cost-saving strategy for commercial poultry farms," *Energy*, vol. 244, p. 123169, 2022.
- [20] P. Oza, P. Sharma, S. Patel, and P. Kumar, "Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey," *Neural Computing and Applications*, pp. 1–22, 2022.
- [21] E. Paquet and F. Soleymani, "Quantumleap: Hybrid quantum neural network for financial predictions," *Expert Systems with Applications*, vol. 195, p. 116583, 2022.
- [22] A. El Bouchti, Y. Tribis, T. Nahhal, and C. Okar, "Forecasting financial risk using quantum neural networks," in *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*. IEEE, 2018, pp. 386–390.
- [23] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, 2017.
- [24] A. Sagingalieva, M. Kordzanganeh, N. Kenbayev, D. Kosichkina, T. Tomashuk, and A. Melnikov, "Hybrid quantum neural network for drug response prediction," *arXiv preprint arXiv:2211.05777*, 2022.
- [25] A. N. Hassan, M. J. Mohsin, and A. H. Khwayyir, "Discrimination of ecg signal based on s-interpolation and quantum neural network," *Annals of Tropical Medicine & Public Health*, vol. 23, no. 10, 2020.
- [26] M. Esposito, G. Uehara, and A. Spanias, "Quantum machine learning for audio classification with applications to healthcare," in *2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2022, pp. 1–4.
- [27] R. Narain, S. Saxena, and A. K. Goyal, "Cardiovascular risk prediction: a comparative study of framingham and quantum neural network based approach," *Patient preference and adherence*, pp. 1259–1270, 2016.
- [28] N. Mathur, J. Landman, Y. Y. Li, M. Strahm, S. Kazdaghi, A. Prakash, and I. Kerenidis, "Medical image classification via quantum neural networks," *arXiv preprint arXiv:2109.01831*, 2021.
- [29] M. J. Umer, J. Amin, M. Sharif, M. A. Anjum, F. Azam, and J. H. Shah, "An integrated framework for covid-19 classification based on classical and quantum transfer learning from a chest radiograph," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 20, p. e6434, 2022.
- [30] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nature Physics*, vol. 15, no. 12, pp. 1273–1278, 2019.
- [31] J. Wu, Z. Tao, and Q. Li, "wpscalable quantum neural networks for classification," in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2022, pp. 38–48.
- [32] T. Hur, L. Kim, and D. K. Park, "Quantum convolutional neural network for classical data classification," *Quantum Machine Intelligence*, vol. 4, no. 1, pp. 1–18, 2022.
- [33] A. Chalumuri, R. Kune, and B. Manoj, "A hybrid classical-quantum approach for multi-class classification," *Quantum Information Processing*, vol. 20, no. 3, pp. 1–19, 2021.
- [34] Z. Abohashima, M. Elhosen, E. H. Houssein, and W. M. Mohamed, "Classification with quantum machine learning: A survey," *arXiv preprint arXiv:2006.12270*, 2020.
- [35] M. Henderson, S. Shakya, S. Pradhan, and T. Cook, "Quantum convolutional neural networks: powering image recognition with quantum circuits," *Quantum Machine Intelligence*, vol. 2, no. 1, pp. 1–9, 2020.
- [36] C. Blank, D. K. Park, J.-K. K. Rhee, and F. Petruccione, "Quantum classifier with tailored quantum kernel," *npj Quantum Information*, vol. 6, no. 1, pp. 1–7, 2020.
- [37] M. Schuld, M. Fingerhuth, and F. Petruccione, "Implementing a distance-based classifier with a quantum interference circuit," *EPL (Europhysics Letters)*, vol. 119, no. 6, p. 60002, 2017.
- [38] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, "Hierarchical quantum classifiers," *npj Quantum Information*, vol. 4, no. 1, pp. 1–8, 2018.
- [39] S. A. Stein, B. Baheri, D. Chen, Y. Mao, Q. Guan, A. Li, S. Xu, and C. Ding, "Quclassi: A hybrid deep neural network architecture based on quantum state fidelity," *Proceedings of Machine Learning and Systems*, vol. 4, pp. 251–264, 2022.
- [40] S. Y. Kilin, "Quantum optics and quantum information," *Optics and Spectroscopy*, vol. 91, no. 3, pp. 325–326, 2001.
- [41] R. Sweke, F. Wilde, J. Meyer, M. Schuld, P. K. Fährmann, B. Meynard-Piganeau, and J. Eisert, "Stochastic gradient descent for hybrid quantum-classical optimization," *Quantum*, vol. 4, p. 314, 2020.
- [42] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, "General parameter-shift rules for quantum gradients," *Quantum*, vol. 6, p. 677, 2022.
- [43] E. Toninelli, B. Ndagano, A. Vallés, B. Sephton, I. Nape, A. Ambrosio, F. Capasso, M. J. Padgett, and A. Forbes, "Concepts in quantum state tomography and classical implementation with intense light: a tutorial," *Advances in Optics and Photonics*, vol. 11, no. 1, pp. 67–134, 2019.
- [44] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 132–149.
- [45] "Open-source quantum development." [Online]. Available: <https://qiskit.org/>
- [46] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.
- [47] "Iris data set." [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/iris>
- [48] D. Bokhan, A. S. Mastiukova, A. S. Boev, D. N. Trubnikov, and A. K. Fedorov, "Multiclass classification using quantum convolutional neural networks with hybrid quantum-classical learning," *arXiv preprint arXiv:2203.15368*, 2022.
- [49] G. Li, Z. Song, and X. Wang, "Vsq: Variational shadow quantum learning for classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 9, 2021, pp. 8357–8365.
- [50] N. A. Nghiem, S. Y.-C. Chen, and T.-C. Wei, "Unified framework for quantum classification," *Physical Review Research*, vol. 3, no. 3, p. 033056, 2021.
- [51] R.-B. Wu, X. Cao, P. Xie, and Y.-x. Liu, "End-to-end quantum machine learning implemented with controlled quantum dynamics," *Physical Review Applied*, vol. 14, no. 6, p. 064020, 2020.
- [52] W. J. Yun, H. Baek, and J. Kim, "Projection valued measure-based quantum machine learning for multi-class classification," *arXiv preprint arXiv:2210.16731*, 2022.
- [53] Z. Yang and X. Zhang, "Entanglement-based quantum deep learning," *New Journal of Physics*, vol. 22, no. 3, p. 033041, 2020.